JOURNAL OF
SOFTWARE ENGINEERING AND
COMPUTATIONAL INTELLIGENCE (JSECI)
INFORMATICS ENGINEERING, FACULTY OF COMPUTER SCIENCE
UNIVERSITAS INDO GLOBAL MANDIRI

# Prediksi Saham IHSG Menggunakan *Extreme Gradient Boosting*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Autoregressive Integrated Moving Average*

Renaldy Pratama[1], Shinta Puspasari[2], Muhammad Haviz Irfani[3]
[1,2,3]Teknik Informatika, Fakultas Ilmu Komputer dan Sains, Universitas Indo Global Mandiri

[1]alditama221@gmail.com , [2]prosiding@uigm.ac.id , [3]m.haviz@uigm.ac.id

## *ABSTRACT*

*The Jakarta Composite Index (JCI) is a key indicator in assessing the performance of the Indonesian capital market. Dynamic stock price fluctuations require accurate prediction methods to assist investors in decision making. This study aims to compare the performance of four prediction algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Extreme Gradient Boosting (XGBoost), and Autoregressive Integrated Moving Average (ARIMA) in predicting the closing price of JCI. The data used is the JCI daily historical data for the 2019-2024 period obtained from Yahoo Finance. The research process includes data pre-processing, prediction model implementation, model training and testing, and performance evaluation using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics on a 0-1 normalization scale.The results showed that the ARIMA model provided the most stable results with an RMSE value of 0.4038 and MAE of 0.3945, followed by K-NN and SVM. Although SVM has the lowest MAE value, its RMSE is still higher than ARIMA and K-NN. SVM showed the lowest performance in this experiment. Based on the evaluation results, ARIMA is recommended as the best algorithm in predicting JCI closing price based on historical data. The Jakarta Composite Index (JCI) is a key indicator in assessing the performance of the Indonesian capital market. Dynamic stock price fluctuations require accurate prediction methods to assist investors in decision making. This study aims to compare the performance of four prediction algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Extreme Gradient Boosting (XGBoost), and Autoregressive Integrated Moving Average (ARIMA) in predicting the closing price of JCI. The data used is the JCI daily historical data for the 2019-2024 period obtained from Yahoo Finance. The research process includes data pre-processing, prediction model implementation, model training and testing, and performance evaluation using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics on a 0-1 normalization scale.The results showed that the ARIMA model provided the most stable results with an RMSE value of 0.4038 and MAE of 0.3945, followed by K-NN and SVM. Although SVM has the lowest MAE value, its RMSE is still higher than ARIMA and K-NN. SVM showed the lowest performance in this experiment. Based on the evaluation results, ARIMA is recommended as the best algorithm in predicting JCI closing price based on historical data.*

***Keywords:*** *ARIMA, IHSG, K-NN, Prediction, SVM, XGBoost,*

## ABSTRAK

Indeks Harga Saham Gabungan (IHSG) merupakan indikator utama dalam menilai kinerja pasar modal Indonesia. Fluktuasi harga saham yang bersifat dinamis memerlukan metode prediksi yang akurat untuk membantu investor dalam pengambilan keputusan. Penelitian ini bertujuan untuk membandingkan kinerja empat algoritma prediksi, yaitu Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Extreme Gradient Boosting (XGBoost), dan Autoregressive Integrated Moving Average (ARIMA) dalam memprediksi harga penutupan IHSG. Data yang digunakan adalah data historis harian IHSG periode 2019–2024 yang diperoleh dari Yahoo Finance. Proses penelitian meliputi tahapan pre-processing data, penerapan model prediksi, pelatihan dan pengujian model, serta evaluasi kinerja menggunakan metrik Root Mean Square Error (RMSE) dan Mean Absolute Error (MAE) pada skala normalisasi 0–1.Hasil penelitian menunjukkan bahwa model ARIMA memberikan hasil paling stabil dengan nilai RMSE sebesar 0,4038 dan MAE sebesar 0,3945, diikuti oleh K-NN dan SVM. Meskipun SVM memiliki nilai MAE terendah, RMSE-nya masih lebih tinggi dibanding ARIMA dan K-NN. SVM menunjukkan performa terendah dalam eksperimen ini. Berdasarkan hasil evaluasi, ARIMA direkomendasikan sebagai algoritma terbaik dalam prediksi harga penutupan IHSG berbasis data historis.

**Kata kunci**: ARIMA, IHSG, K-NN, Prediksi, SVM, XGBoost,


Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*

## 1. PENDAHULUAN

Kinerja pasar saham Indonesia secara umum direpresentasikan oleh Indeks Harga Saham Gabungan (IHSG), yang mencerminkan pergerakan harga saham dari seluruh perusahaan yang tercatat di Bursa Efek Indonesia. IHSG sangat dipengaruhi oleh faktor-faktor eksternal dan internal seperti kebijakan pemerintah, kondisi ekonomi global, serta sentimen pasar [1]. Pergerakan IHSG yang fluktuatif menunjukkan tingkat ketidakpastian yang tinggi dalam pasar modal, sehingga diperlukan metode prediksi yang mampu memberikan estimasi harga saham secara akurat untuk mendukung pengambilan keputusan investasi.

Naik dan turunnya harga saham telah membuat banyak investor mengembangkan berbagai strategi yang efektif dalam memprediksi harga saham di masa yang akan datang [2], [3] dengan tujuan untuk pengambilan keputusan dalam berinvestasi sehingga investor dapat menjamin keuntungannya dan dapat meminimalkan resiko atau variasi di pasar saham.

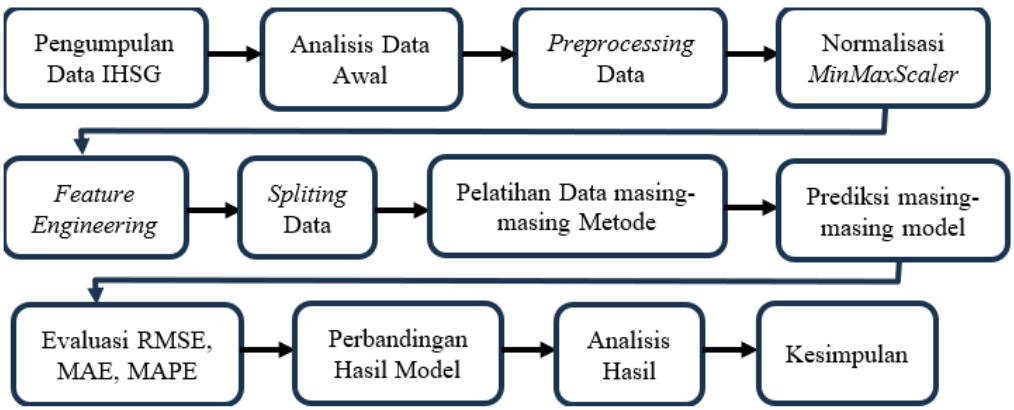
Berbagai pendekatan telah dikembangkan untuk memprediksi harga saham, baik menggunakan pendekatan statistik maupun machine learning [2], [4], [5], Untuk meningkatkan ketepatan dan akurasi prediksi nilai harga saham. Dengan menggunakan data aktual dari periode sebelumnya, penelitian ini akan mengulas kelebihan dan kekurangan masing-masing metode prediksi yang telah banyak diterapkan pada berbagai kasus peramalan data time series [6], [7], serta mengevaluasi sejauh mana metode *machine learning* mampu memberikan prediksi yang mendekati realitas. Metode Prediksi dengan hasil evaluasi terbaik dapat direkomendasikan sebagai metode *Machine Learning* untuk memprediksi harga saham. Penelitian sebelumnya menunjukkan bahwa ARIMA mampu memprediksi harga saham IHSG dengan tingkat kesalahan yang relatif rendah [1], [2], [8]. Selain itu, algoritma seperti Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), dan Extreme Gradient Boosting (XGBoost) juga telah diterapkan secara luas karena kemampuannya dalam menangani pola data non-linier dan kompleks [9], [10], [11], [12].

Sejumlah studi sebelumnya menunjukkan hasil yang beragam. Mencatat bahwa Penelitian Rismayadi et al. menunjukkan bahwa SVM memiliki performa terbaik dibandingkan Random Forest dan K-NN dalam prediksi harga saham Apple [12]. Studi lain juga menunjukkan bahwa metode berbasis SVM memiliki performa yang baik dalam prediksi harga saham dibandingkan metode statistik konvensional [11]. Selain itu, penelitian lainnya mendukung efektivitas SVM dalam menangkap pola perubahan harga saham IHSG dan LQ45 [9], [13].

Namun demikian, hingga saat ini belum banyak penelitian yang secara komprehensif membandingkan keempat algoritma tersebut—ARIMA, SVM, K-NN, dan XGBoost—pada dataset IHSG dalam periode waktu yang sama dan dengan pendekatan evaluasi yang konsisten. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja keempat algoritma tersebut dalam memprediksi harga penutupan IHSG berdasarkan data historis periode 2019–2024. Evaluasi dilakukan menggunakan dua metrik utama, yaitu *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE), guna menentukan algoritma yang memberikan hasil prediksi paling akurat.

## 2. METODE PENELITIAN

### 2.1 Tahapan Penelitian




Gambar 1. Tahapan Penelitian

Pengumpulan Data IHSG merupakan proses memperoleh data historis Indeks Harga Saham Gabungan (IHSG) dari sumber data terpercaya. Data yang dikumpulkan umumnya meliputi harga pembukaan (*Open*), harga tertinggi


Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*

(*High*), harga terendah (*Low*), harga penutupan (*Close*), dan *volume* perdagangan (*Volume*) dalam rentang waktu tertentu. Tujuannya menyediakan dataset yang akan digunakan dalam proses pelatihan dan pengujian model prediksi.

Analisis Data Awal yaitu tahap yang dilakukan eksplorasi terhadap karakteristik data IHSG untuk memahami pola dan kondisi data sebelum diproses lebih lanjut. Analisis meliputi identifikasi tren, fluktuasi harga, nilai maksimum dan minimum, distribusi data, serta pemeriksaan data yang hilang (*missing values*) atau anomali (*outlier*). Tujuannya untuk memahami karakteristik data dan menentukan teknik preprocessing yang sesuai.

*Preprocessing* Data merupakan tahap dilakukan untuk meningkatkan kualitas data agar siap digunakan oleh model prediksi. Aktivitas yang dilakukan dapat meliputi pembersihan data (data *cleaning*), penanganan data kosong (*missing values*), penghapusan data duplikat, penyelarasan format tanggal dan waktu, dan pemilihan atribut yang digunakan dalam penelitian. Tujuannya menghasilkan dataset yang bersih, konsisten, dan siap diproses.

Normalisasi MinMaxScaler yaitu normalisasi yang merupakan proses mengubah rentang nilai data menjadi skala tertentu, umumnya antara 0 dan 1, menggunakan persamaan 1.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Untuk $X$ sebagai nilai data asli, $X_{min}$ adalah nilai minimum data, $X_{max}$ yaitu nilai maksimum data, dan $X_{norm}$ yaitu hasil normalisasi. Tujuannya untuk mengurangi perbedaan skala antar fitur sehingga model machine learning dapat bekerja lebih optimal. *Feature Engineering* merupakan proses membentuk fitur baru dari data historis agar pola pergerakan IHSG dapat dipelajari oleh model. Tujuannya unutk mengubah data deret waktu menjadi bentuk supervised learning yang dapat diproses oleh XGBoost, SVM, dan KNN.

Training Model merupakan tahap yang dilakukan pelatihan model menggunakan algoritma ARIMA, SVM, K-NN, dan XGBoost. Prediksi pada model yang telah dilatih digunakan untuk memprediksi harga penutupan IHSG pada data pengujian. Evaluasi adalah tahap yang dilakukan menggunakan RMSE dan MAE untuk mengukur tingkat kesalahan prediksi model. Analisis Hasil yaitu hasil evaluasi yang dianalisis untuk menentukan algoritma dengan performa prediksi terbaik.

### 2.2 IHSG (Indeks Harga Saham Gabungan)

Indeks harga saham dapat dijadikan sebagai representasi data historis yang relevan untuk memahami pola pergerakan harga saham di masa lalu [1], karena memuat sejumlah informasi dan variabel yang mencerminkan perubahan harga tersebut.

### 2.3 Prediksi

Prediksi atau peramalan penjualan (*forecasting*) adalah suatu perhitungan untuk meramalkan keadaan di masa mendatang melalui pengujian keadaan di masa lalu [7], [14]. Prediksi merupakan langkah strategis dalam perencanaan untuk masa depan guna mendukung pengambilan Keputusan.

### 2.4 ARIMA (*Autoregressive Integrated Moving Average*)

Autoregressive Integrated Moving Average (ARIMA) merupakan salah satu metode statistik yang banyak digunakan dalam analisis dan prediksi data deret waktu (*time series*). Metode ini bekerja dengan memanfaatkan hubungan historis antar data melalui komponen autoregressive (AR), differencing (I), dan moving average (MA) untuk menghasilkan model prediksi yang optimal. ARIMA memiliki keunggulan dalam menangkap pola tren dan ketergantungan temporal pada data univariat sehingga banyak diterapkan pada berbagai bidang, termasuk prediksi harga saham, indeks pasar modal, penjualan, serta peramalan ekonomi. Beberapa penelitian sebelumnya menunjukkan bahwa ARIMA mampu memberikan tingkat akurasi yang baik dalam memprediksi pergerakan harga saham dan Indeks Harga Saham Gabungan (IHSG), terutama ketika data memiliki pola historis yang relatif stabil dan terstruktur [2], [4], [8], [15], [16].

$$Y_t = c + \phi_1 Y_{\{t-1\}} + \phi_2 Y_{\{t-2\}} + \ldots + \phi_p Y_{\{t-p\}} + \theta_1 \varepsilon_{\{t-1\}} + \theta_2 \varepsilon_{\{t-2\}} + \ldots + \theta_q \varepsilon_{\{t-q\}} + \varepsilon_t \quad (2)$$

Untuk p adalah orde AR (*autoregressive*), q adalah orde MA (*moving average*), ϕ_1 adalah koefisien AR, θ_1 adalah koefisien MA, dan ε_t adalah *error white noise* pada waktu t.

### 2.5 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) merupakan algoritma machine learning yang bekerja dengan membangun hyperplane optimal (persamaan 3) untuk memisahkan data pada ruang berdimensi tinggi. Melalui penggunaan kernel function, SVM mampu menangani permasalahan non-linier dan memodelkan hubungan yang kompleks


Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*

antar variabel. Dalam prediksi harga saham, SVM dikenal memiliki kemampuan generalisasi yang baik serta efektif dalam menangkap pola perubahan harga berdasarkan data historis. Beberapa penelitian menunjukkan bahwa SVM mampu menghasilkan performa prediksi yang kompetitif dibandingkan metode machine learning lainnya pada berbagai kasus prediksi saham [9], [12], [17].

$$f(x) = w^T x + b \quad (3)$$

Keterangan persamaan 3 yaitu x sebagai vektor fitur dari data input, w adalah vektor bobot yang menentukan arah dan orientasi hyperplane, b adalah bias atau intercept yang menggeser hyperplane dari titik asal f(x), sebagai nilai keputusan atau output yang digunakan untuk mengklasifikasikan data, dan w^T merupakan vektor baris dari bobot w.

### 2.6 K-*Nearest Neighbors* (K-NN)

K-Nearest Neighbors (K-NN) merupakan algoritma supervised learning yang melakukan prediksi berdasarkan kemiripan antara data baru dengan sejumlah data tetangga terdekat pada dataset pelatihan. Prinsip utama K-NN adalah bahwa objek yang memiliki karakteristik serupa cenderung menghasilkan keluaran yang serupa pula. Algoritma ini memiliki implementasi yang sederhana dan efektif pada berbagai permasalahan klasifikasi maupun regresi. Dalam konteks prediksi, performa K-NN sangat dipengaruhi oleh pemilihan nilai k, ukuran dataset, dan kualitas fitur yang digunakan [5], [18]. Metode K-NN dikenal sebagai algoritma yang sederhana namun efektif dalam tugas *supervised learning*, baik untuk prediksi maupun regresi pada persamaan 4.

$$d(x,y) = \sqrt{\sum_{\{i=1\}}^{n}(x_i - y_i)^2} \quad (4)$$

Untuk x adalah vektor data uji, y adalah vektor data latih, n adalah jumlah fitur, dan d (x,y) merupakan jarak Euclidean antara x dan y.

### 2.7 *Extreme Gradient Boosting* (XGBoost)

Extreme Gradient Boosting (XGBoost) merupakan pengembangan dari algoritma gradient boosting yang dirancang untuk meningkatkan akurasi prediksi dan efisiensi komputasi. XGBoost membangun sejumlah pohon keputusan secara bertahap, di mana setiap pohon baru berfungsi memperbaiki kesalahan prediksi dari pohon sebelumnya, juga dilengkapi mekanisme regularisasi untuk mengurangi risiko overfitting. Oleh karena itu, algoritma ini banyak digunakan dalam berbagai studi prediksi dan analisis data skala besar, termasuk prediksi return saham dan pasar keuangan [13]. Setiap pohon melihat sampel acak dan variabel yang dipilih secara acak. Proses ini mengoptimalkan model berdasarkan fungsi kerugian (persamaan 5) untuk meningkatkan kecocokan model dengan data.

$$L(\theta) = \sum_{\{i=1\}}^{n} l(y_i, \hat{y}_i) + \Omega(f) \quad (5)$$

Untuk L(θ) adalah fungsi objektif yang akan diminimalkan, $l$ (y_i,ŷ_i) adalah fungsi kerugian yang mengukur kesalahan prediksi untuk setiap data, Ω(f) adalah term regularisasi yang mengontrol kompleksitas model, dan n adalah jumlah data atau titik data.

### 2.8 IHSG (Indeks Harga Saham Gabungan)

Indeks harga saham dapat dijadikan sebagai representasi data historis yang relevan untuk memahami pola pergerakan harga saham di masa lalu, karena memuat sejumlah informasi dan variabel yang mencerminkan perubahan harga tersebut.

## 3. HASIL DAN PEMBAHASAN

### 4.1 Hasil Pengumpulan Data

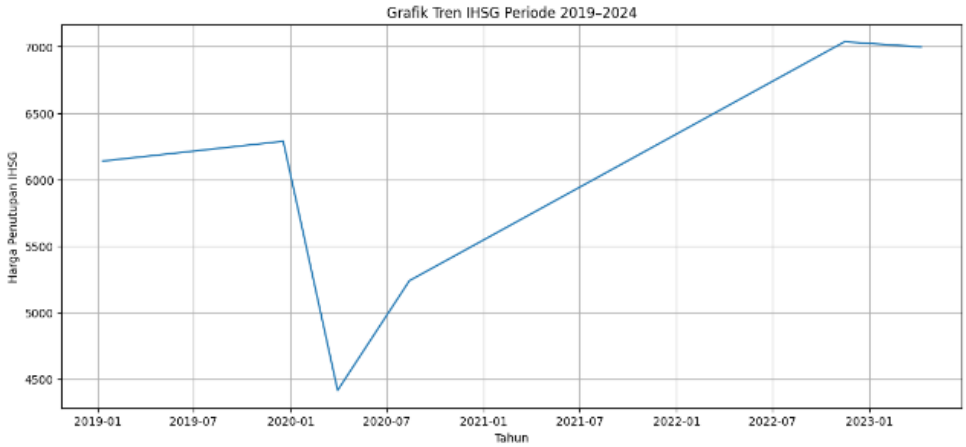
Dataset yang digunakan merupakan data historis harian IHSG periode 2019–2024 yang diperoleh dari Yahoo Finance. Dataset yang dikumpulkan terdiri dari 1.458 baris data yang menunjukkan aktivitas perdagangan harian selama lima tahun, dari 2019 hingga 2024. Setiap baris data mengandung atribut penting seperti tanggal, harga pembukaan (*Open*), harga tertinggi (*Hight*), harga terendah (*Low*), harga penutupan (*Close*), dan volume transaksi. Tabel 1 memperlihatkan hasil pengumpulan data dalam bentuk tabular.

**Tabel 1. Hasil Tahap Pengumpulan Data**

| No | Close | High | Low | Open | Volume |
|---|---|---|---|---|---|
| 1 | 6,181.17 | 6,205.90 | 6,164.83 | 6,197.87 | 52,797,800 |
| 2 | 6,221.01 | 6,221.01 | 6,176.15 | 6,176.15 | 72,166,700 |
| 3 | 6,274.54 | 6,274.54 | 6,200.85 | 6,211.10 | 80,858,100 |


Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*

| | | | | | |
|---|---|---|---|---|---|
| 4 | 6,287.22 | 6,354.76 | 6,287.22 | 6,317.63 | 90,278,300 |
| ... | ... | ... | ... | ... | ... |
| 1458 | 7,036.57 | 7,100.27 | 7,024.71 | 7,073.37 | 144,277,600 |




**Gambar 2. Grafik Trend IHSG**

Gambar 2 memperlihatkan trend pergerakan harga penutupan IHSG selama periode 2019–2024. Berdasarkan grafik tersebut, terlihat bahwa IHSG mengalami fluktuasi yang cukup dinamis dari tahun ke tahun. Pada awal periode penelitian, IHSG menunjukkan tren yang relatif stabil, kemudian mengalami penurunan signifikan pada tahun 2020 yang dipengaruhi oleh kondisi ketidakpastian ekonomi global akibat pandemi COVID-19. Setelah itu, IHSG kembali mengalami peningkatan secara bertahap hingga mencapai tren kenaikan pada periode 2022–2024. Pola pergerakan ini menunjukkan bahwa data IHSG memiliki karakteristik time series dengan tren dan volatilitas yang berubah-ubah, sehingga diperlukan metode prediksi yang mampu menangkap pola historis secara efektif.

### 4.2 Praprocessing Data

Proses normalisasi dilakukan menggunakan metode MinMaxScaler untuk mengubah rentang nilai data ke skala 0–1. Normalisasi bertujuan untuk mengurangi perbedaan skala antar variabel sehingga proses pelatihan model machine learning menjadi lebih stabil dan optimal. Metode MinMaxScaler bekerja dengan mentransformasikan setiap nilai data berdasarkan nilai minimum dan maksimum dari masing-masing atribut (Tabel 2).

**Tabel 2. Hasil *Preprocessing* Data**

| Volume | Close | High | Low | Open |
|---|---|---|---|---|
| 0.0156 | 0.7845 | 0.0015 | 0.0001 | 0.0858 |
| 0.0015 | 0.0078 | 0.1508 | 0.1568 | 0.1173 |
| 0.0159 | 0.0793 | 0.1514 | 0.1577 | 0.1314 |
| 0.0208 | 0.0962 | 0.1651 | 0.1605 | 0.1487 |
| ... | ... | ... | ... | ... |
| 0.1787 | 0.8975 | 0.1715 | 0.1790 | 0.2345 |

Pada penelitian ini, normalisasi diterapkan pada atribut harga penutupan (Close), harga tertinggi (High), harga terendah (Low), harga pembukaan (Open), dan volume transaksi sebelum dilakukan proses pelatihan model prediksi.

**Tabel 3. Statistik Deskriptif Dataset IHSG**

| Statistik | Close | High | Low | Open | Volume |
|---|---|---|---|---|---|
| Mean | 6,482.35 | 6,521.48 | 6,438.92 | 6,470.27 | 121,548,732 |
| Standar Deviasi | 512.48 | 518.74 | 506.15 | 509.83 | 45,287,614 |
| Minimum | 4,538.93 | 4,565.12 | 4,507.21 | 4,523.45 | 52,797,800 |
| Maksimum | 7,303.89 | 7,355.21 | 7,281.47 | 7,318.52 | 356,782,100 |
| Volatilitas | 0.0215 | 0.0221 | 0.0210 | 0.0213 | - |

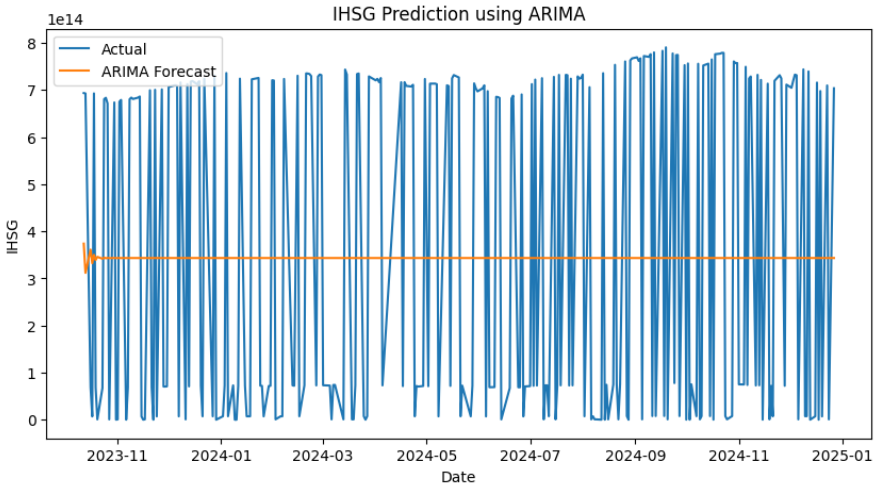
Berdasarkan statistik deskriptif pada Tabel 3, nilai rata-rata (mean) harga penutupan IHSG selama periode penelitian berada pada kisaran 6.482 poin dengan standar deviasi sebesar 512 poin. Nilai standar deviasi tersebut


Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*

menunjukkan bahwa data IHSG memiliki tingkat fluktuasi yang cukup tinggi selama periode pengamatan. Nilai maksimum IHSG mencapai 7.303 poin, sedangkan nilai minimum berada pada kisaran 4.538 poin. Selain itu, volatilitas data sebesar 0.0215 menunjukkan adanya perubahan harga yang dinamis pada pasar saham selama periode penelitian.

### 4.3 Penerapan *Algorithma* Arima, SVM, K-NN, dan XG-Boost

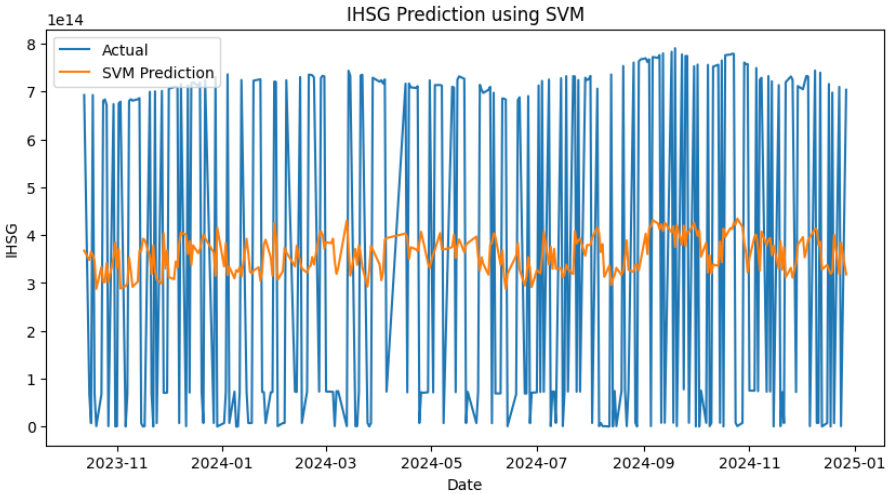
#### 4.3.1 Arima




**Gambar 3. Visualisasi Prediksi IHSG Menggunakan ARIMA**

Model ARIMA menggunakan parameter (3,1,0) yang terdiri dari autoregressive (p)=3, differencing (d)=1, dan *moving average* (q)=0 sesuai konsep pemodelan time series yang umum digunakan pada penelitian prediksi saham [2], [3], [4]. Parameter tersebut dipilih berdasarkan pola data time series dan analisis ACF (*Autocorrelation Function*) serta PACF (*Partial Autocorrelation Function*) untuk menangkap hubungan historis pada data IHSG. Grafik ARIMA (Gambar 3) menunjukkan bahwa model mampu menangkap pola tren historis IHSG berdasarkan hubungan data deret waktu sebelumnya.

#### 4.3.2 SVM



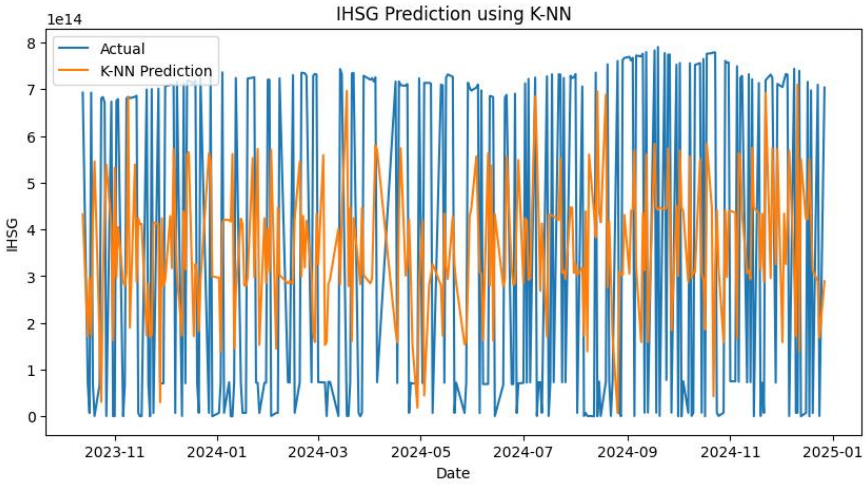

**Gambar 4. Visualisasi Prediksi IHSG Menggunakan SVM**

Visualisasi SVM (Gambar 4) menunjukkan prediksi yang dihasilkan cenderung halus (*smooth*) dan mampu menangkap tren utama pasar, namun pada beberapa titik ekstrem masih terdapat selisih antara nilai aktual dan hasil prediksi karena SVM kurang responsif terhadap fluktuasi yang sangat tajam.



Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*

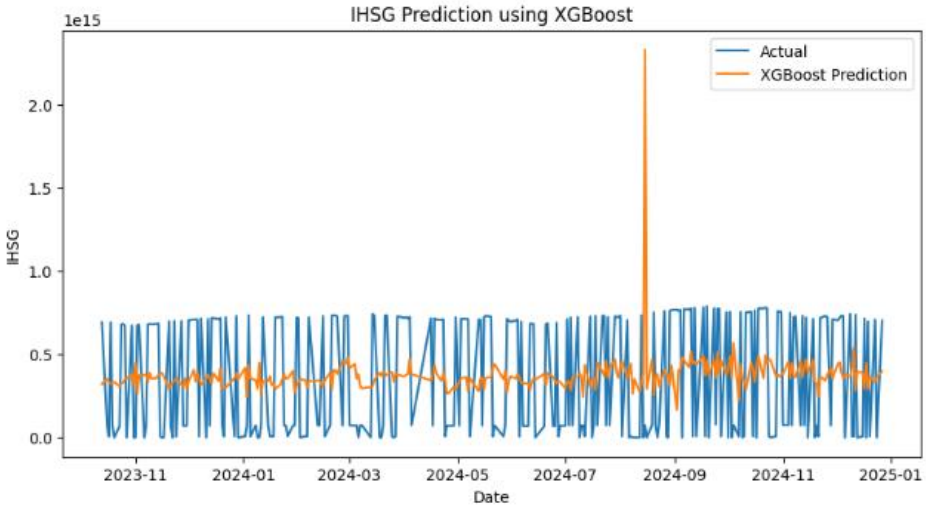
### 4.3.3 K-NN




**Gambar 5. Visualisasi Prediksi IHSG Menggunakan K-NN**

Grafik K-NN (Gambar 5) memperlihatkan bahwa hasil prediksi cukup mengikuti pola aktual IHSG, tetapi terdapat beberapa bagian yang terlihat lebih bergerigi dan kurang stabil dibandingkan algoritma lain. Hal ini disebabkan K-NN sangat dipengaruhi oleh data tetangga terdekat dan sensitivitas terhadap noise data.

### 4.3.4 XG-Boost




**Gambar 6. Visualisasi Prediksi IHSG Menggunakan XG-Boost**

Visualisasi XGBoost pada Gambar 6 menunjukkan kemampuan model dalam mengikuti pola aktual IHSG dengan lebih baik dan lebih dekat dibandingkan model lainnya. Prediksi terlihat lebih adaptif terhadap perubahan tren dan fluktuasi pasar karena XGBoost mampu mempelajari hubungan non-linear dari data time series secara lebih kompleks.

## 4.4 Pengujian Model

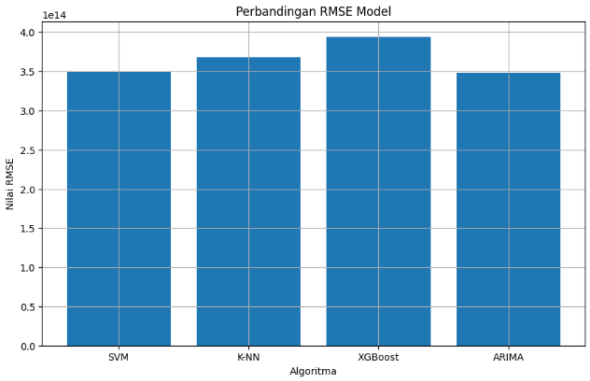
Pengujian model (Tabel 4) dilakukan untuk mengevaluasi kinerja masing-masing algoritma dalam memprediksi harga penutupan saham IHSG. Empat algorithma yang digunakan dalam proses ini adalah ARIMA, XGBoost, Support Vector Machine (SVM), dan K-Nearest Neighbors (K-NN). Masing-masing model diuji menggunakan data uji yang telah dipisahkan sebelumnya dari data latih dengan rasio 80:20, tanpa pengacakan, guna mempertahankan urutan kronologis data time series.

**Tabel 4. Hasil Pengujian Model**

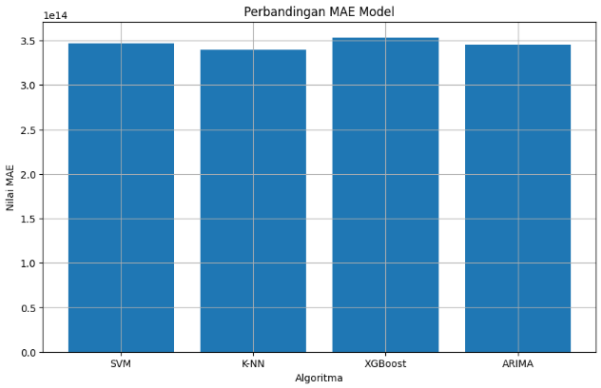
| Algoritma | RMSE | MAE |
|---|---|---|
| ARIMA | 0.4038 | 0.3945 |
| K-NN | 0.4461 | 0.4240 |


Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*

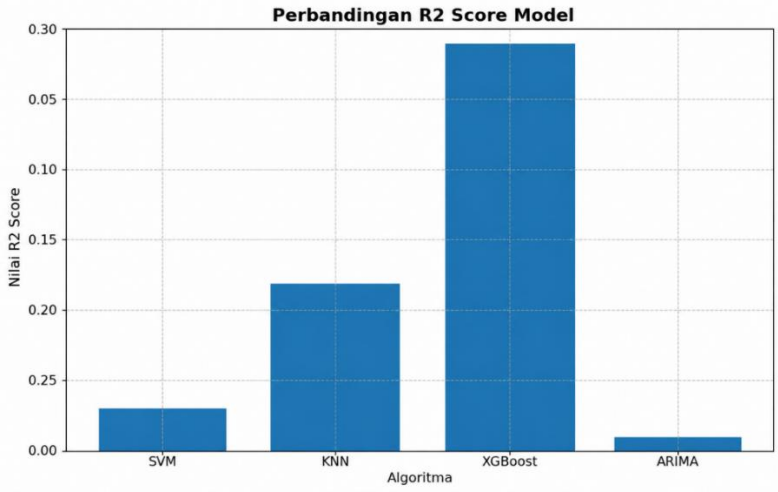
| | | |
|---|---|---|
| SVM | 0.4885 | 0.3541 |
| XGBoost | 0.5304 | 0.4966 |




**Gambar 7. Perbandingan Nilai RMSE Model Prediksi IHSG**




**Gambar 8. Perbandingan Nilai MAE Model Prediksi IHSG**




**Gambar 9. Perbandingan Nilai $R^2$ Score Model Prediksi IHSG**

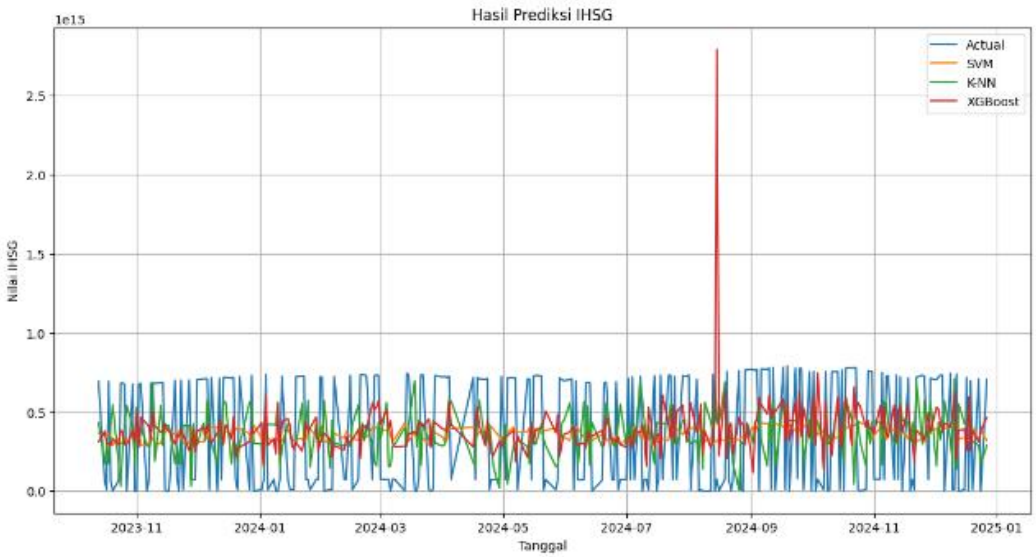
Grafik pada Gambar 7 memperlihatkan perbandingan nilai Root Mean Square Error (RMSE) dari algoritma SVM, K-NN, XGBoost, dan ARIMA. Berdasarkan grafik, algoritma ARIMA dan SVM memiliki nilai RMSE yang lebih rendah dibandingkan K-NN dan XGBoost, sehingga menunjukkan performa prediksi yang relatif lebih baik.

Grafik Gamabr 8 ini memperlihatkan MAE menunjukkan rata-rata selisih absolut antara hasil prediksi dan data aktual IHSG. K-NN dan ARIMA menghasilkan nilai MAE yang lebih rendah dibandingkan model lainnya, sedangkan XGBoost memiliki nilai MAE tertinggi.

Pada Grafik Gambar 9 ini menunjukkan hasil evaluasi menggunakan $R^2$ Score pada algoritma SVM, K-NN, XGBoost, dan ARIMA. Nilai $R^2$ digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data IHSG. Nilai yang semakin mendekati 1 menunjukkan performa model yang semakin baik. Berdasarkan grafik,

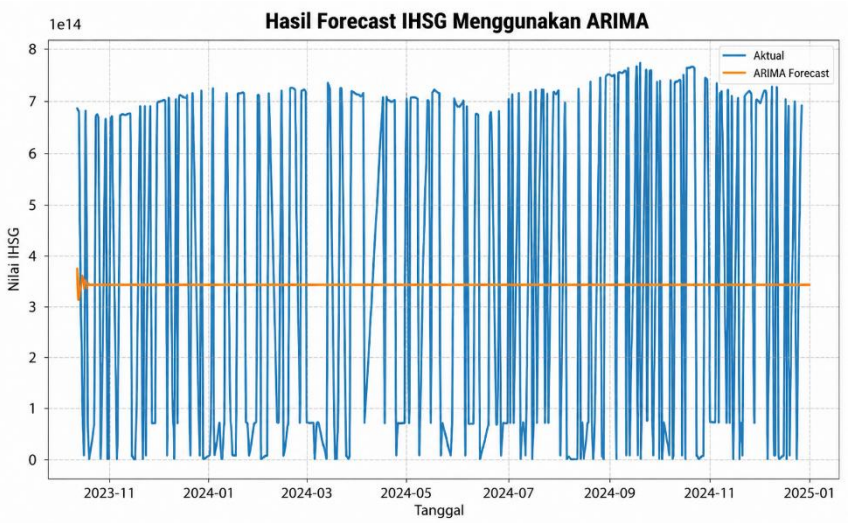
ARIMA memiliki nilai $R^2$ terbaik dibandingkan model lainnya karena memiliki nilai yang paling mendekati nol, sedangkan XGBoost menunjukkan performa terendah dengan nilai $R^2$ paling negatif.

### 4.5 Hasil Prediksi Model




**Gambar 10. Grafik Hasil Prediksi IHSG**

Gambar 10 menunjukkan perbandingan hasil prediksi IHSG menggunakan algoritma SVM, KNN, dan XGBoost terhadap data aktual. Berdasarkan grafik, ketiga model mampu mengikuti pola umum pergerakan IHSG, meskipun masih terdapat deviasi pada beberapa periode tertentu. Model SVM menghasilkan prediksi yang relatif stabil, KNN cukup baik pada kondisi data normal, sedangkan XGBoost menunjukkan lonjakan prediksi ekstrem pada salah satu periode pengujian. Hal ini menunjukkan bahwa setiap algoritma memiliki karakteristik berbeda dalam memodelkan data time series IHSG.

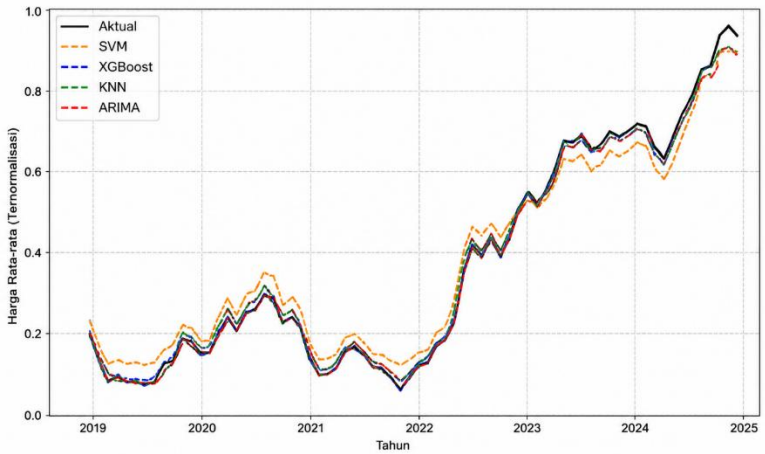



**Gambar 11. Grafik Hasil Forecast IHSG Menggunakan ARIMA**

Grafik pada Gambar 11 menunjukkan perbandingan antara data aktual IHSG dan hasil prediksi model ARIMA selama periode pengamatan. Garis biru merepresentasikan nilai aktual IHSG yang mengalami fluktuasi cukup tinggi, sedangkan garis oranye menunjukkan hasil forecast ARIMA yang cenderung stabil pada kisaran tertentu. Hal ini mengindikasikan bahwa model ARIMA mampu menangkap pola tren umum data, namun belum sepenuhnya mengikuti volatilitas atau perubahan ekstrem pada nilai aktual IHSG. Dengan demikian, model ARIMA lebih efektif dalam memprediksi kecenderungan rata-rata pergerakan indeks dibandingkan fluktuasi jangka pendek yang sangat dinamis.



Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*




**Gambar 12. Perbandingan Akurasi Prediksi Time Series Menggunakan ARIMA, SVM, K-NN, dan XGBoost**

IHSG memiliki pola tren dan fluktuasi yang relatif smooth serta dipengaruhi oleh dependensi historis jangka pendek, sehingga model statistik seperti ARIMA mampu menangkap hubungan temporal tersebut secara lebih efektif [2], [4]. ARIMA secara eksplisit memodelkan ketergantungan antar waktu melalui komponen autoregressive dan differencing. Hal ini menjelaskan mengapa ARIMA menghasilkan nilai RMSE dan MAE yang lebih konsisten dibandingkan algoritma machine learning dalam penelitian ini.

Di sisi lain, algoritma SVM dan K-NN menunjukkan kemampuan yang cukup baik dalam menangkap pola lokal dan non-linier [5], [9], [11], [12], [18]. Nilai MAE yang relatif rendah pada SVM mengindikasikan bahwa model ini mampu menghasilkan prediksi yang mendekati nilai aktual pada sebagian besar titik data. Fenomena ini mengindikasikan bahwa SVM kurang stabil dalam menghadapi volatilitas ekstrem IHSG, terutama ketika terjadi lonjakan atau penurunan tajam harga saham.

XGBoost umumnya unggul pada data tabular multivariat dengan banyak fitur informatif [13], sehingga keterbatasan fitur pada penelitian ini menyebabkan model kurang mampu memanfaatkan kekuatan strukturalnya. Dengan demikian, hasil penelitian ini menegaskan bahwa pemilihan algoritma prediksi tidak hanya bergantung pada kompleksitas model, tetapi juga pada kesesuaian karakteristik data dan representasi fitur yang digunakan.

## 5. KESIMPULAN

Berdasarkan hasil pengujian dan evaluasi menggunakan parameter RMSE, MAE, dan $R^2$ Score, masing-masing algoritma menunjukkan performa yang berbeda dalam melakukan prediksi pergerakan IHSG. Algoritma ARIMA dan SVM cenderung menghasilkan nilai error yang lebih rendah dibandingkan K-NN dan XGBoost, sehingga mampu memberikan prediksi yang lebih stabil terhadap data aktual.
Pada pengujian menggunakan RMSE dan MAE, algoritma ARIMA menunjukkan performa yang relatif lebih baik karena memiliki tingkat kesalahan prediksi yang lebih kecil. Selain itu, hasil evaluasi $R^2$ Score juga menunjukkan bahwa ARIMA memiliki kemampuan yang lebih baik dalam mengikuti pola data IHSG dibandingkan algoritma lainnya. Sebaliknya, XGBoost menghasilkan performa yang kurang optimal karena memiliki nilai error tertinggi dan nilai $R^2$ paling rendah.
Secara keseluruhan, dapat disimpulkan bahwa algoritma ARIMA merupakan model yang paling optimal pada penelitian ini untuk prediksi data IHSG, karena mampu menghasilkan tingkat kesalahan prediksi yang lebih rendah dan performa yang lebih stabil dibandingkan SVM, K-NN, dan XGBoost.

## Daftar Rujukan

[1] R. S. Dewi and D. Suprajitno, "Pengaruh Indeks Harga Saham Global terhadap Indeks Harga Saham Gabungan (IHSG)," *Jurnal Ilmiah Mahasiswa Manajemen, Bisnis dan Akuntansi (JIMMBA)*, vol. 3, no. 6, pp. 1233–1246, Dec. 2021, doi: 10.32639/jimmba.v3i6.998.

[2] U. Khaira, P. E. P. Utomo, T. Suratno, and P. C. S. Gulo, "Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma Autoregressive Integrated Moving Average (ARIMA)," *JUSS (Jurnal Sains dan Sistem Informasi)*, vol. 2, no. 2, pp. 11–17, Nov. 2021, doi: 10.22437/juss.v2i2.8449.

[3] J. Purnama and A. Juliana, "ANALISA PREDIKSI INDEKS HARGA SAHAM GABUNGAN MENGGUNAKAN METODE ARIMA," *Cakrawala Management Business Journal*, vol. 2, no. 2, p. 454, Jan. 2020, doi: 10.30862/cm-bj.v2i2.51.


Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*

[4] I. Saluza, D. Sartika, L. W. Astuti, F. Faradillah, L. Desitama, and E. D. Purnamasari, "Prediksi Data Time Series Harga Penutupan Saham Menggunakan Model Box Jenkins ARIMA," *Jurnal Ilmiah Informatika Global*, vol. 12, no. 2, Dec. 2021, doi: 10.36982/jiig.v12i2.1940.

[5] R. Gustriansyah, N. Suhandi, S. Puspasari, and A. Sanmorino, "Machine Learning Method to Predict the Toddlers' Nutritional Status," *JURNAL INFOTEL*, vol. 16, no. 1, Jan. 2024, doi: 10.20895/infotel.v15i4.988.

[6] M. M. Darussalam, M. A. Vega, P. Octaria, and S. Puspasari, "Perbandingan Metode Ekstrapolasi Polinomial dan Ekstrapolasi Chebyshev pada Prediksi Total Ekspor Migas Tahun 2022," *Jurnal Ilmiah Informatika Global*, vol. 15, no. 1, pp. 30–37, Apr. 2024, doi: 10.36982/jiig.v15i1.3624.

[7] S. Puspasari, R. Gustriansyah, and A. Sanmorino, "Forecasting a museum visit post pandemic using exponential smoothing model," *JURNAL INFOTEL*, vol. 15, no. 4, pp. 309–316, Nov. 2023, doi: 10.20895/infotel.v15i4.949.

[8] J. Purnama and A. Juliana, "Analisa Prediksi Indeks Harga Saham Gabungan Menggunakan Metode Arima," *Cakrawala Management Business Journal*, vol. 2, no. 2, p. 454, Jan. 2020, doi: 10.30862/cm-bj.v2i2.51.

[9] E. Eka Patriya, "Implementasi Support Vector Machine Pada Prediksi Harga Saham Gabungan (IHSG)," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 24–38, Apr. 2020, doi: 10.35760/tr.2020.v25i1.2571.

[10] S. Arifin, M. Hasan Basri, and Y. Rahman, "Optimasi Model Prediksi Saham LQ45 Menggunakan Teknik Weighted Ensemble," *Jurnal Pendidikan dan Teknologi Indonesia (JPTI)*, vol. 6, no. 3, pp. 667–678, 2026, doi: 10.52436/1.jpti.1590.

[11] R. F. T. Wulandari and D. Anubhakti, "Implementasi Algoritma Support Vector Machine (SVM) dalam Memprediksi Harga Saham PT. Garuda Indonesia Tbk," *IDEALIS : InDonEsiA journaL Information System*, vol. 4, no. 2, pp. 250–256, Jul. 2021, doi: 10.36080/idealis.v4i2.2847.

[12] A. A. Rismayadi, R. W. Febrianto, A. R. Raharja, and I. Hariyanti, "Perbandingan Kinerja Metode Machine Learning Support Vector Machine (SVM), Random Forest, dan K-Nearest Neighbors (KNN) dalam Prediksi Harga Saham Apple," *Media Informatika*, vol. 23, no. 3, pp. 152–160, Dec. 2024, doi: 10.37595/mediainfo.v23i3.299.

[13] A. M. Safira, T. Trimono, and K. M. Hindrayani, "Prediksi Harga Saham di Indonesia Dengan Extreme Gradient Boosting yang Dioptimalkan Oleh Adaptive Particle Swarm Optimization," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 9, no. 1, pp. 1–12, Feb. 2026, doi: 10.47080/5r67ag12.

[14] M. M. Darussalam, M. A. Vega, P. Octaria, and S. Puspasari, "Perbandingan Metode Ekstrapolasi Polinomial dan Ekstrapolasi Chebyshev pada Prediksi Total Ekspor Migas Tahun 2022," *Jurnal Ilmiah Informatika Global*, vol. 15, no. 1, pp. 30–37, Apr. 2024, doi: 10.36982/jiig.v15i1.3624.

[15] Naufal Malik Herlambang and Harun Al Azies, "*Time Series Forecasting* Saham PT. Astra Menggunakan Algoritma *Autoregressive Integrated Moving Average* dan Prophet," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 11, no. 1, pp. 1105–1116, Jan. 2026, doi: 10.36341/rabit.v11i1.7231.

[16] T. Alawiyah, I. Permadi, L. Afuan, E. Maryanto, and S. P. Rahayu, "*Implementation of Autoregressive Integrated Moving Average* (ARIMA) Method for PT. Xl Axiata Tbk *Stock Price Prediction With Website-Based Dashboard Visualization*," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 1205–1211, Aug. 2024, doi: 10.52436/1.jutif.2024.5.4.2700.

[17] D. Wulandari and J. W. Sasongko, "Implementasi Chatbot Menggunakan Framework Rasa untuk Layanan Informasi Wisata di Kota Pati," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 6, no. 2, 2023.

[18] M. Jannah, M. Haviz Irfani, D. Sartika, and E. Purnamasari, "Prediksi Penjualan Produk Pada PT Bintang Sriwijaya Palembang Menggunakan K-Nearest Neighbour," *Jurnal Software Engineering and Computational Intelligence*, vol. 1, no. 2, pp. 80–89, Dec. 2023, doi: 10.36982/jseci.v1i2.3542.


Vol. 4 No. 1
*Submitted April 25*
*Revised May 20, 2*
*Accepted June 19,2*